Cluster Analysis for the Evaluation of Temporal and Spatial Variations in Water Quality of Wadi Henifa (Riyadh – Saudi Arabia) Case-Study

Dr. Mohab Amin
Department of Civil Engineering
King Saud University

Introduction

- Wadis (valleys) play a major role in assimilation or carrying off the municipal and industrial wastewater and run-off from agricultural land.
- Seasonal variations in precipitation, surface run-off, interflow, groundwater flow and pumped in and outflows have a strong effect on wadi discharge and subsequently on the concentration of pollutants in wadi water.
- Therefore the effective and long term management of stream requires a fundamental understanding of major factors affecting the change in water quality parameters.
- However, due to spatial and temporal variations in surface water quality, a monitoring program that produces a large and complex data set is required to provide a representative and reliable estimation of the water quality.

Introduction

- In view of the spatial and temporal variations in hydrochemistry of water streams, regular monitoring programs are required for reliable estimates of the water quality. This results in a huge and complex data matrix comprised of a large number of physico-chemical parameters (Chapman, 1992), which are often difficult to interpret; drawing meaningful conclusions (Dixon and Chiswell, 1996).
- The size and complexity of the resulting multivariate datasets, formulating overall conclusions regarding the importance of sources of contamination and environmental patterns are often difficult when the constituents are considered one at a time. Thus, multivariate statistical methods can be a useful adjunct compared to traditional univariate methods of data analysis (Helena et al., 2000).

Introduction

- In view of the spatial and temporal variations in hydrochemistry of water streams, regular monitoring programs are required for reliable estimates of the water quality. This results in a huge and complex data matrix comprised of a large number of physico-chemical parameters (Chapman, 1992), which are often difficult to interpret; drawing meaningful conclusions (Dixon and Chiswell, 1996).
- The size and complexity of the resulting multivariate datasets, formulating overall conclusions regarding the importance of sources of contamination and environmental patterns are often difficult when the constituents are considered one at a time. Thus, multivariate statistical methods can be a useful adjunct compared to traditional univariate methods of data analysis (Helena et al., 2000).

Introduction

- Wadi Hanifa
  - On the Najd plateau of central Saudi Arabia, Wadi Hanifa runs southeast for around 120 kilometers (75 mi) before losing itself on the fringes of the Rub' al-Khali, or Empty Quarter. Fed by more than 40 tributaries, this great watercourse has a catchment area covering much of the eastern Najd, more than 4500 square kilometers (1740 sq mi), across what was historically known as al-Yamamah.
  - Approximately 1,000,000 cubic meters of water (dry weather flow) continually flows out of Riyadh (the capital city of Saudi Arabia) each day into Wadi Hanifa from various side wadis and channels.

Wadi Hanifa Location (Modified from Alhamid, et al., 2007).

Introduction

- Water quality of Wadi Hanifa system in Saudi Arabia is regularly monitored at several sites for a large number of physico-chemical, bacteriological and hydrological parameters with an outcome of tremendous databases of high complexity.
- Such monitoring programs involve huge financial inputs. Thus, there is a need to optimize the monitoring networks, number of water-quality parameters, reducing these to representative ones without losing useful information. The multivariate statistical techniques and exploratory data analysis are the appropriate tools for a meaningful data reduction and interpretation of multi-constituent chemical and physical measurements (Massart et al., 1988).

Introduction

- In the present study, a multivariate statistical technique (cluster analysis) is applied to evaluate both the spatial and temporal variations in water-quality data matrix of the Wadi Hanifa, Riyadh, Saudi Arabia, which were generated under 1 year monitoring program.
- The research aims at identifying the most informative sampling sites and studying the influence of natural and artificial sources affecting the variation of water quality parameters in Wadi Hanifa, on both temporal and spatial levels

Monitored Parameters

- In order to represent the water quality of the Wadi system, accounting for stream and inputs from drains that have impact on downstream water quality, the monitoring and sampling plan were designed to cover a wide range of determinants at specific sites.
- Under the water-quality monitoring plan of Wadi Hanfia, samples were collected each month at three points across the Wadi width at several sites.
- Water temperature was measured on site using mercury thermometer. All other parameters were determined in laboratory following the standard protocols (APHA, 1992). The samples were analyzed for 23 parameters, which are pH, specific conductivity (Sp. Cond.), alkalinity (Alk), Dissolved Oxygen (DO), suspended solids (SS), total dissolved solids (TDS), volatile suspended solids (VSS), total dissolved solids (TDS), dissolved oxygen (DO), 5-days biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammonia nitrogen (NH_3-N), nitrate nitrogen (NO_3-N), nitrite nitrogen (NO_2-N), total kjeldahl nitrogen (TKN), total nitrogen (TN), total phosphorus (T-P), Ortho-phosphate (Ortho-PO_4^{3-}), Total organic carbon (TOC), Bicarbonate (HCO_3^-), turbidity (Trub) and salinity (Sal). All the water quality parameters are expressed in mg/L, except pH, EC (mS/cm), Total Colifrom (CFU/100 ml), Fecal Colifrom (CFU/100 ml) and temperature (^oC).

Monitoring Sites

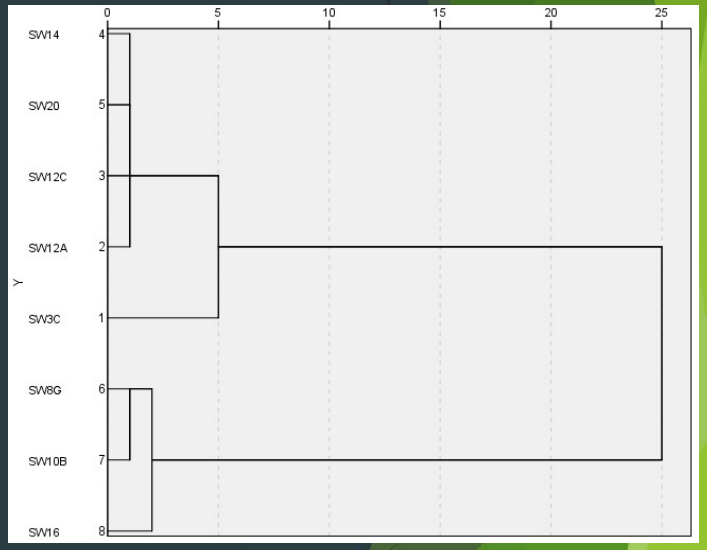
- Eight sites were selected to cover the full length of the wet part of Wadi Hanifa as shown in Figure 2. Their names are SW3C (site 1), SW12A (site 2), SW12C (site 3), SW14 (site 4), SW20 (site 5), SW8G (site 6), SW10B (site 7), and SW16 (site 8). The sites location are chosen as four sites before the bioremediation facility and four sites after the facility. However, site 5 is before the connection between Batha Channel, which carries water from Manfoha wastewater treatment plant and the Wadi.

Cluster Analysis

- Cluster analysis is a group of techniques used mainly to group objects similar to each other in a cluster (category or class).
- Objects with different characteristics are directed, by the analysis mechanism, into different clusters.
- The analysis results in exposing the hidden structure of the data under consideration.
- Cluster analysis does not provide an explanation or interpretation of the described data structure.
- Hierarchical clustering is the most common approach wherein clusters are formed sequentially. The most similar pair of objects is joined first in one cluster, and the rest of objects are added to the cluster according to their level of similarity. The Euclidean distance is usually used to measure the similarity between two samples. The distance is measured repetitively, and the pair of two clusters with the smallest distance, at each step, are grouped. (Wunderlin et al., 2001, Vega et al., 1998, MH Sayadi, et al., 2014).

Results and Discussions
Temporal Study

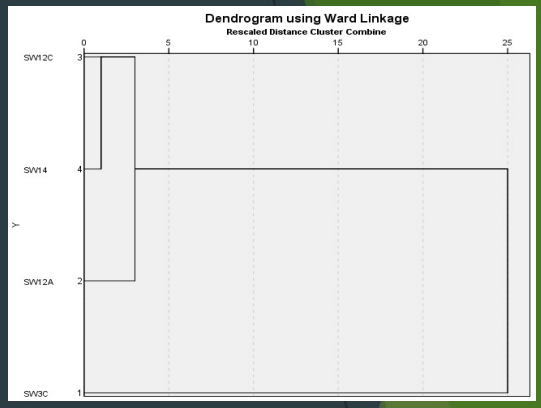
- Cluster analysis on the temporal level produced a dendrogram.
- All eight sampling sites on the Wadi were grouped into two statistically significant clusters at (D_{link}/D_{max}) x100 < 60 (less than 15 out of 25 on the figure shown below) as mentioned by Shrestha and Kazama 2007.
- **The clustering procedure, on the temporal level, generated two groups of sites in a very definite way, as the sites in these groups have similar background source types, where sites from 1 to 5 are before the connection between batha channel and the Wadi.**
- Cluster 1 which includes Sites 1 – 5 are highly related, and cluster 2 which includes Sites 6 – 8 are highly related, however the two clusters were not related.
- **This implies that for rapid assessment of water quality, only one site in each cluster may serve as good in temporal assessment of the water quality for the whole network.**
- There are other reports (Wunderlin et al., 2001; Simeonov et al., 2003), where this approach has successfully applied in water-quality programs.



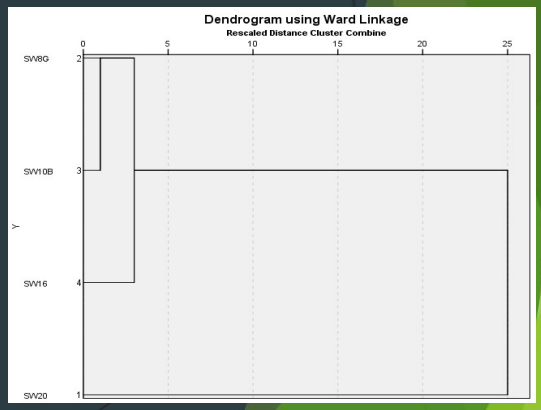
Dendrogram showing clustering of sampling on Wadi Hanifa on temporal level.

Results and Discussions
Spatial Study

- Cluster analysis on the spatial level produced a dendrogram. The Cluster analyses for the first group (Sites 1 to 4) showed that sampling stations from 2 to 4 ((D_{link}/D_{max}) x100 < 15) were statistically significant cluster. Also the cluster analyses for the second group (Sites 5 to 8) showed that sampling stations from 6 to 8 ((D_{link}/D_{max}) x100 < 15) were statistically significant cluster. Only one site in each group from the clusters shown by the cluster analyses may serve as good in spatial assessment of the water quality.
- SW 3C is not related to the rest of sites because it is located upstream of the wadi and it did not get contaminated by the side wadis and channels. SW 20 is located before batha channel that is the reason SW 8G, SW 10B and SW 16 are one cluster and SW 20 is not related to the rest.



a. Group 1



b. Group 2

Dendrogram showing clustering of sampling on Wadi Hanifa on Spatial level.

Conclusions

- Cluster analyses are effective means of manipulating, interpreting and representing data concerning the surface water quality. Eight monitoring sites were selected to cover the full length of the wet part of Wadi Hanifa. Water-quality monitoring of the Wadi Hanfia was regularly conducted over a period of one year. In this case study, two groups of data both for temporal and spatial evaluations have been selected. On both the temporal and spatial levels only one site in each group from the clusters may serve as good in the assessment of the water quality. The clustering procedure, on the temporal level, generated two groups of sites in a very definite way, as the sites in these groups have similar background source types, where sites from 1 to 5 are before the connection between batha channel and the Wadi.
- Cluster analysis on the spatial level showed that SW3C is not related to the rest of sites because it is located upstream of the wadi and it did not get contaminated by the side wadis and channels. From the results of CA, it appears that the effect of station on the data decreases as the station being evaluated is getting far from the batha channel. It has been also conclude that the bioremediation facility works inconsistently during different seasons. During raining seasons the bioremediation facility does not work efficiently due to the increase of flow above its capacity.